Nonreversible MCMC for discontinuous targets and disconnected spaces with boundaries

Jere Koskela

Stochastic analysis of the neutron transport equation
17 September 2021

EPSRC
Engineering and Physical Sciences Research Council

Warwick Statistics

Outline

- The zig-zag process
- The zig-zag process on a hybrid state space
- The coalescent
- A geometric embedding of coalescent trees
- Zig-zag for the coalescent
- One slide on neutrons

The zig-zag process[1]

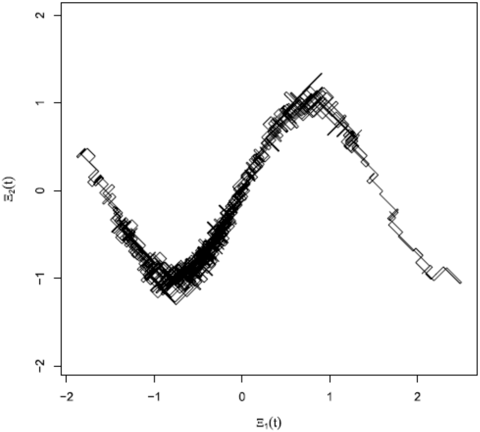
1. Set $(x, v) \in \mathbb{R}^n \times \{-1, 1\}^n$.
2. Set $t \leftarrow 0$.
3. While $t < T$ do
   3.1 Sample $Y \sim \text{Exp}(1)$.
   3.2 Set ρ such that $\sum_{i=1}^n \int_t^{t+\rho} \lambda_i(x + sv, v)\mathrm{d}s = Y$.
   3.3 Set $x \leftarrow x + \rho v$.
   3.4 Set $t \leftarrow t + \rho$.
   3.5 Sample $I \sim \text{Categorical}(\lambda_1(x, v), \ldots, \lambda_n(x, v))$.
   3.6 Set $v_I \leftarrow -v_I$.

- Between switches, x moves with constant velocity v.
- Target density π is invariant if

$$\lambda_i(x, v) = v_i \partial_i \log \pi(x) \vee 0.$$

[1] J Bierkens, P Fearnhead and G Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data, Ann Stat 47(3):1288–1320, 2019.

The zig-zag process



(d) 2D S-shaped density

[1] J Bierkens, P Fearnhead and G Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data, Ann Stat 47(3):1288–1320, 2019.

Advantages and disadvantages of zig-zag

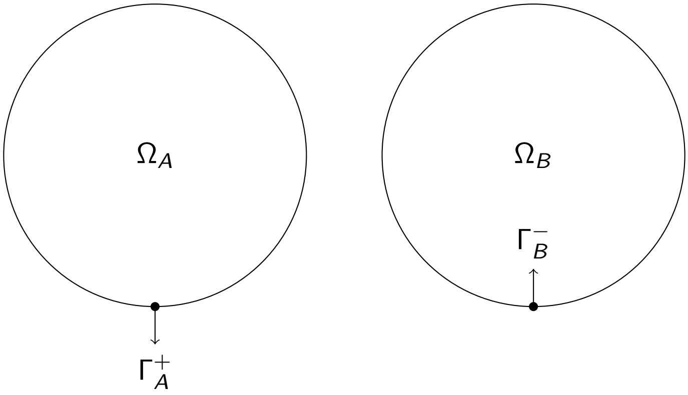
+ Avoids reversible backtracking.
+ Rejection-free.
+ Direction of travel guided by shape of target.
− Construction only makes sense if target density is differentiable. . .
− and if the state space is one connected set.

The zig-zag process on a hybrid state space

- Target $\pi(m, x)$ on $\mathbb{F} \times \Omega \subseteq \mathbb{N} \times \mathbb{R}^n$.
- For each $m \in \mathbb{F}$ create Ω_m, a copy of Ω.
- Let $\Gamma_m^{\pm}$ be respective exit and entrance boundaries into Ω_m (which can be artificially introduced if necessary), and suppose Q is a Markov kernel from $\Gamma^+ = \cup_{m \in \mathcal{F}} \Gamma_m^+$ to $\Gamma^- := \cup_{m \in \mathcal{F}} \Gamma_m^-$.
- While in some Ω_m^o, let the process undergo zig-zag dynamics with target $\pi(m, \cdot)$.
- When an exit boundary is hit, the state jumps with law Q and re-enters a different copy.

The zig-zag process on a hybrid state space

Take $\mathbb{F} = \{A, B\}$ and $\Omega = B(0, 1)$.



Piecewise deterministic Markov processes[2]

The generalised zig-zag process (m_t, x_t, v_t) is a PDMP with generator

$$Lf(m,x,v) := v \cdot \nabla_x f(m,x,v) + \sum_{i=1}^{d} \lambda_i(m,x,v)[f(m,x,F_i v) - f(m,x,v)]$$

whose domain $D(L)$ consists of measurable functions f satisfying

$$f(m,x,v) = \int_{(j,y,w)\in\Gamma^-} f(j,y,w) Q(m,x,v;j,\mathrm{d}y,w)$$

for any $(m,x,v) \in \Gamma^+$, and some boundedness & continuity conditions.

[2] M Davis. Markov models and optimization, Chapman & Hall, 1993. Especially chapters 2 and 3.

Theorem[3]

Suppose the initial condition has a density, that $\mathbb{F}$ is finite, that $\pi(m, \cdot) \in C^1(\Omega_m^o)$ for each $m \in \mathbb{F}$, that $\pi > 0$ on $\cup_{m \in \mathbb{F}} \Omega_m^o$, and

1. $\mathbb{E}_{(m,x,v)}[\#\text{jumps by time } t] < \infty$ for any (m, x, v) and $t > 0$,
2. for each $(x, v) \in \Gamma_m^+$ and $(y, w) \in \Gamma_j^-$,

$$\pi(m, x) Q(m, x, v; j, y, w) = \pi(j, y) Q(j, y, -w; m, x, -v),$$

3. for each $(x, v) \in \Gamma_m^+$,

$$\int_{(j,y,w) \in \Gamma^-} (w \cdot n(j, y)) Q(m, x, v; j, \mathrm{d}y, w) = -v \cdot n(m, x),$$

   where $n(m, x)$ is the unit outward normal at $x \in \partial \Omega_m$,
4. and that Q is such that jumps facing corners are a null event.

Then π is a stationary distribution of (m_t, x_t, v_t).

[3] J Koskela. Zig-zag sampling for discrete structures and non-reversible phylogenetic MCMC, arXiv:2004.08807.

Sketch proof

- The various technical assumptions guarantee that $C_b^1 \cap D(L)$ separates measures on the domain of the zig-zag process.
- The aim is to show $\mathbb{E}_\pi[Lf(m,x,v)] = 0$ for each $f \in C_b^1 \cap D(L)$.

$$\mathbb{E}_\pi\left[v \cdot \nabla_x f(m,x,v) + \sum_{i=1}^{d} \lambda_i(m,x,v)[f(m,x,F_i v) - f(m,x,v)]\right]$$
$$= \mathbb{E}_\pi[v \cdot \{\nabla_x f(m,x,v) - f(m,x,v)\nabla_x \log \pi(m,x)\}]$$
$$= \mathbb{E}_\pi[v \cdot \nabla_x f(m,x,v) - v \cdot \nabla_x f(m,x,v) - \text{Boundary terms}].$$

- The defining boundary property of the domain $D(L)$ and the two conditions on π and Q are exactly what you need to show that boundary terms vanish.

Ergodicity

If $\mathbb{F}$ is finite and the process restricted to a single domain Ω_m with target $\pi(m, \cdot)$ is ergodic for each $m \in \mathbb{F}$, then the full process is ergodic provided

$$\int_{(x,v)\in\Gamma_m^+} \int_{(y,v)\in\Gamma_j^-} Q(m, x, v; j, \mathrm{d}y, w)\pi(m, x) > 0$$

for enough ordered pairs $(m, j) \in \mathbb{F}^2$ to form a tour.

Some criteria for ergodicity of a single-domain process are known[4].

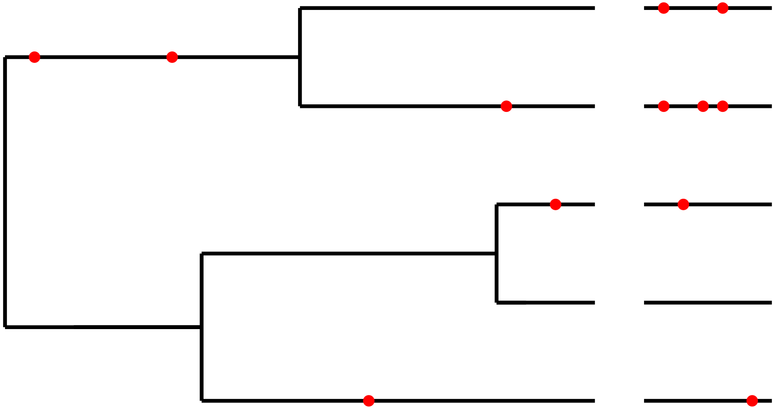
[4] J Bierkens, G Roberts and P-A Zitt. Ergodicity of the zig-zag process, Ann Appl Probab 29(4):2266–2301, 2019.

The coalescent[5]

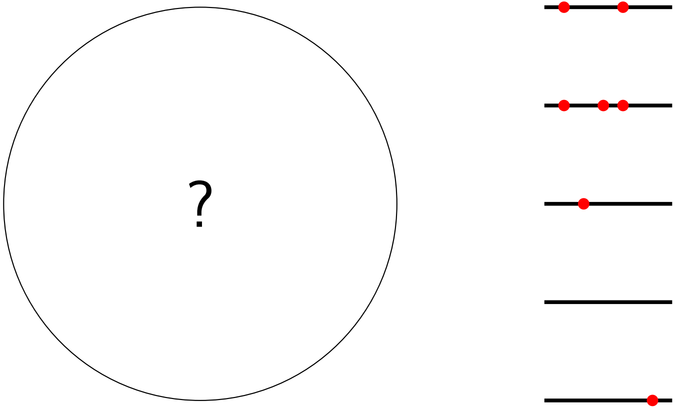
- $(\Pi_t)_{t\geq 0}$ with $\Pi_0 = \{\{1\}, \ldots, \{n\}\}$.
- Holding times $T_k \sim \text{Exp}\left(\binom{k}{2}\right)$.
- At jump times, two randomly chosen blocks merge.
- $i \sim j$ in $\Pi_t \Leftrightarrow i$ and j have a common ancestor at time t.
- Mutations as a Poisson point process with rate $\theta/2$ on the branches of the tree.

[5] J F C Kingman. The coalescent, Stoch Proc Appl 13(3):235–248, 1982.

The coalescent



The coalescent as missing data



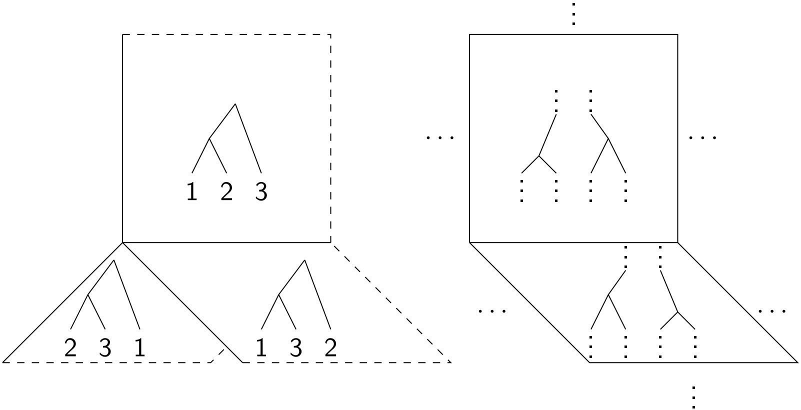
$$P(D|\theta) = \int_A P(D|A,\theta)P(A|\theta)dA$$

A geometric embedding of coalescent trees[6]

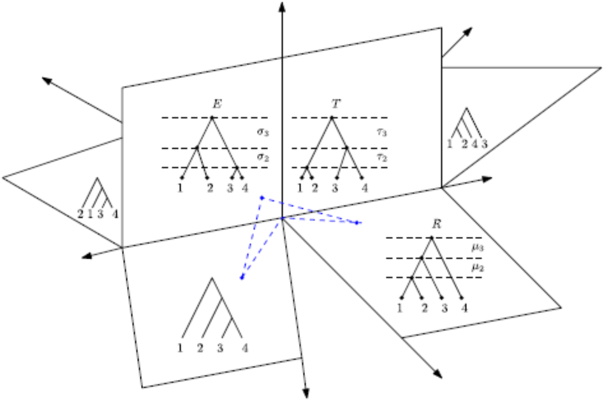
- Encode a tree as $(E_n, \mathbf{t}_n)$.
- $\mathbf{t}_n := (t_1, \ldots, t_{n-1}) \in \mathbb{R}_+^{n-1}$ are the times between successive mergers.
- E_n is the *ranked topology*: an ordered set of nodes specifying the exact merger events.
- τ-space: separate copy of $\mathbb{R}_+^{n-1}$ for each distinct E_n.
- Glue common faces of separate orthants together to form a connected space.
- $\#E_n = \frac{(n-1)!n!}{2^{n-1}}$.

[6] A Gavryushkin and A J Drummond. The space of ultrametric phylogenetic trees, J Theor Biol 403:197–208, 2016.

The three leaf τ-space



One third of the four leaf τ-space



[6] A Gavryushkin and A J Drummond. The space of ultrametric phylogenetic trees, J Theor Biol 403:197–208, 2016.

The infinite sites coalescent posterior

- Regard E_n as a set of edges.
- Given n observed sequences and a putative ancestral tree $(E_n, \mathbf{t}_n)$, for an edge $\gamma \in E_n$, define:
  - $m_\gamma :=$ number of mutations on γ.
  - $t_i \in \gamma$ if γ spans the ith interval.
  - $\ell_\gamma := \sum_{i: t_i \in \gamma} t_i$ be the branch length.

$$\pi(\theta, E_n, \mathbf{t}_n) \propto \left\{ \prod_{\gamma \in E_n} \frac{1}{m_\gamma!} \left(\frac{\theta \ell_\gamma}{2} \right)^{m_\gamma} \right\} \times \exp\left(-\sum_{i=1}^{n-1} \frac{(n+1-i)(n+\theta-i)}{2} t_i \right).$$

Flip rates

$$\lambda_i(E_n, \mathbf{t}_n, \theta; v) := v_i \left(\frac{(n+1-i)(n+\theta-i)}{2} - \sum_{\gamma \in E_n : t_i \in \gamma} \frac{m_\gamma}{\ell_\gamma} \right) \vee 0,$$

$$\lambda_\theta(E_n, \mathbf{t}_n, \theta; v) := v_\theta \left(\sum_{i=1}^{n-1} \frac{n+1-i}{2} t_i - \frac{1}{\theta} \sum_{\gamma \in E_n} m_\gamma \right) \vee 0.$$

Important: $v_i = \pm 1 / \binom{n+1-i}{2}$.

Ranked topology crossings

- Exit and entrance boundaries := faces at which $t_i = 0$ or $\theta = 0$.
- Define the jump kernel Q via the orthants which intersect at a face.
- π is continuous at the boundaries of ranked topologies.
- Crossing into uniformly chosen adjacent topology when a boundary is hit satisfies the conditions on Q.
- So does reflecting at the $t_1 = 0$ and $\theta = 0$ boundaries.

Simulation study[7]

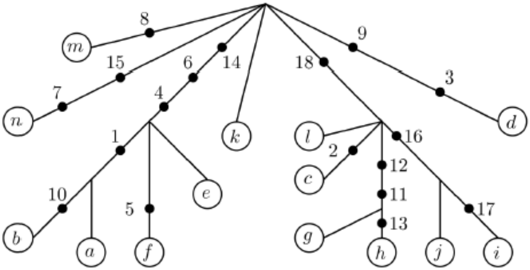


Figure 3.
Perfect phylogeny of the Griffiths and Tavaré (1994) data set.

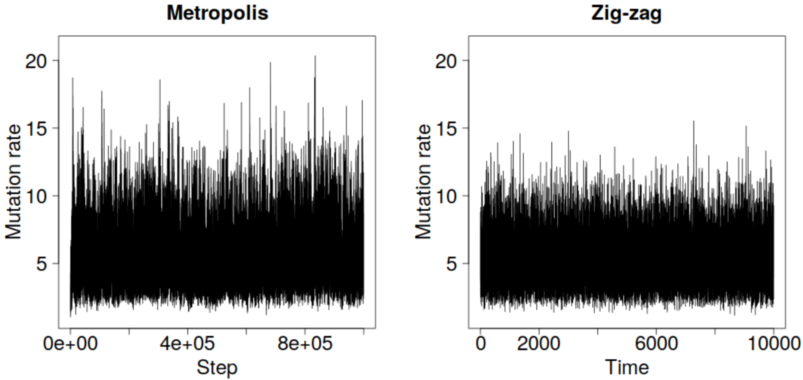
---

[8] A Hobolth, M K Uyenoyama and C Wiuf. Importance sampling for the infinite sites model, Stat Appl Genet Mol Biol 7(1) Article 32, 2008.

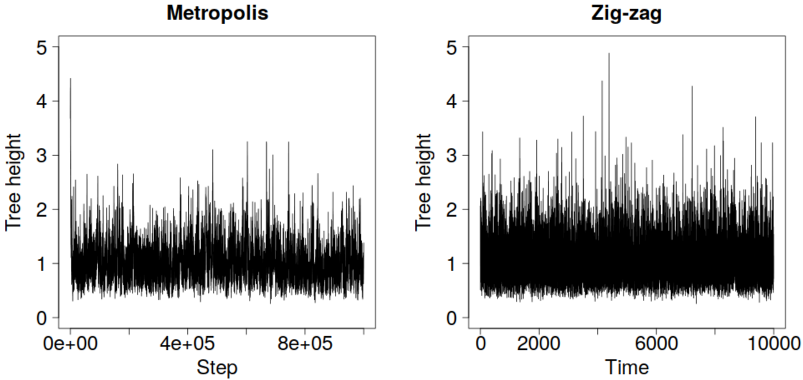
[8] R C Griffiths and S Tavaré. Ancestral inference in population genetics, Stat Sci 9:307–319, 1994.

[8] R H Ward, B L Frazier, K Dew and S Pääbo. Extensive mitochondrial diversity within a single Amerindian tribe, Proc Natl Acad Sci USA 88:8720–8724, 1991.

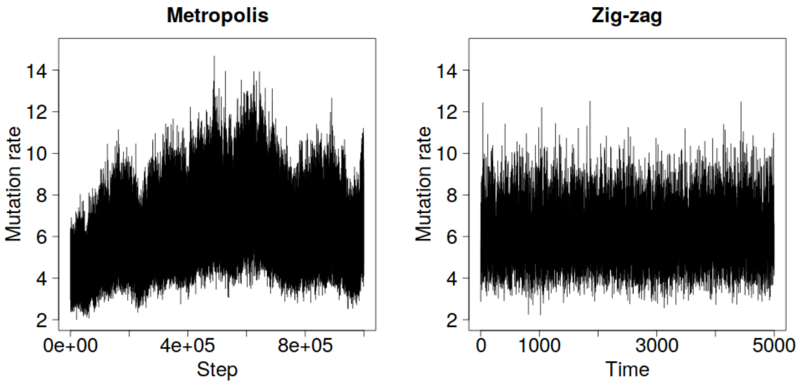
Mutation rate



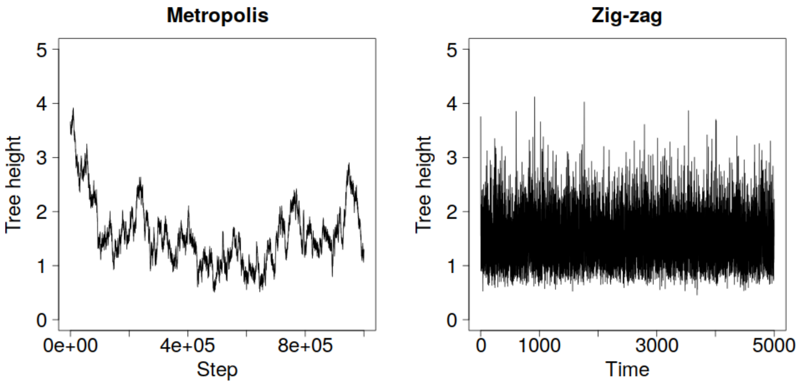
Tree height


Metropolis
Tree height
Step
0e+00
4e+05
8e+05
0
1
2
3
4
5
Zig-zag
Tree height
Time
0
2000
6000
10000
0
1
2
3
4
5

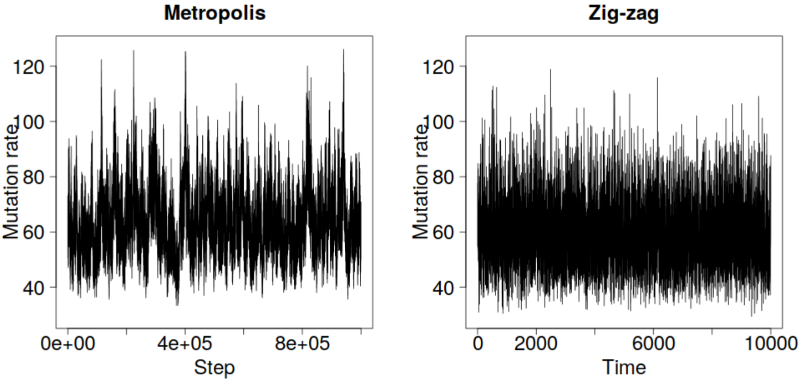
Mutation rate, sample size = 550



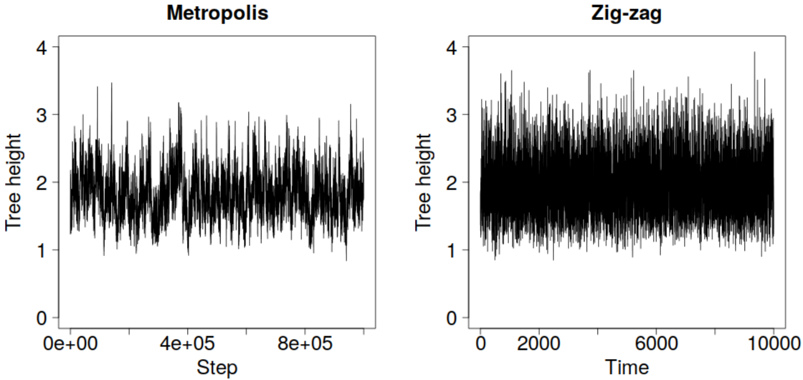
Tree height, sample size = 550



Mutation rate, $\theta = 55$



Tree height, $\theta = 55$



What about a nuclear reactor?